Serial No.: ~~10~~09/745,795
Docket No.: 28951.4056

IN THE CLAIMS:

1. (Currently Amended) A character string dividing system for segmenting a character string into a plurality of words, comprising:

input means for receiving documents to be processed for word division;

document data storing means serving as a document database for storing the received documents to be division processed;

character joint probability calculating means for calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in said document database, in which calculation of said character joint probability is performed based only on the information involved in said document database without referring to any dictionary;

probability table storing means for storing a table of calculated character joint probabilities;

character string dividing means for segmenting an objective character string into a plurality of words with reference to said table of calculated character joint probabilities, without relying on any dictionary; and

output means for outputting a division result of said objective character string.

2. (Currently Amended) A character string dividing method for segmenting a

character string into a plurality of words, said method comprising the steps of:

calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a document database storing document to be processed for word division, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary; and

segmenting an objective character string into a plurality of words with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability, without relying on any dictionary.

3. (Currently Amended) A character string dividing method for segmenting a character string into a plurality of words, said method comprising the steps of:

calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database to be processed for word division, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary, said character joint probability being calculated as an appearance probability of a specific character string appearing immediately before a specific character, said specific character string including a former one of said two neighboring characters as a tail thereof and said specific character being a latter one of said two neighboring characters; and

segmenting an objective character string into a plurality of words with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability, without relying on any dictionary.

4. (Currently Amended) A character string dividing method for segmenting a character string into a plurality of words, said method comprising the steps of:

statistically calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database to be processed for word division, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary, said character joint probability being calculated as an appearance probability of a first character string appearing immediately before a second character string, said first character string including a former one of said two neighboring characters as a tail thereof and said second character string including a latter one of said two neighboring characters as a head thereof; and segmenting an objective character string into a plurality of words with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability, without relying on any dictionary.

5. (Previously Presented) The character string dividing method in accordance with claim 4, wherein said character joint probability of two neighboring characters is calculated based on a first probability of said first character string appearing immediately before said latter one of said two neighboring characters and also based on a second probability of said second character string appearing immediately after said former one of said two neighboring characters.

6. (Currently Amended) A character string dividing method for segmenting a character string into a plurality of words, said method comprising the steps of:

calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database to be processed for word division and prepared for learning purpose in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary; and

segmenting an objective character string into a plurality of words with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability, without relying on any dictionary,

wherein, when said objective character string involves a sequence of characters not involved in said document database, a character joint probability of any two neighboring characters not appearing in said database is estimated based on said calculated character joint probabilities for the neighboring characters stored in said document database.

7. (Previously Presented) The character string dividing method in accordance with claim 2, wherein said division point of said objective character string is determined based on a comparison between the character joint probability and a threshold, and said threshold is determined with reference to an average word length of resultant words.

8. (Previously Presented) The character string dividing method in accordance with claim 2, wherein a changing point of character type is a prospective division point of said objective character string.

9. (Cancelled)

10. (Currently Amended) A character string dividing system for segmenting a

character string into a plurality of words, comprising:

input means for receiving a document to be processed for word division;

document data storing means serving as a document database for storing a received document to be division processed;

character joint probability calculating means for calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in said document database, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary;

probability table storing means for storing a table of calculated character joint probabilities;

word dictionary storing means for storing a word dictionary prepared or produced beforehand;

division pattern producing means for producing a plurality of candidates for a division pattern of an objective character string with reference to information of said word dictionary;

correct pattern selecting means for selecting a correct division pattern from said plurality of candidates with reference to said table of character joint probabilities; and

output means for outputting said selected correct division pattern as a division result of said objective character string.

11. (Currently Amended) A character string dividing method for segmenting a character string into a plurality of words, said method comprising:

calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database to be processed for word division, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary;

storing calculated character joint probabilities; and

segmenting an objective character string into a plurality of words with reference to a word dictionary,

wherein, when there are a plurality of candidates for a division pattern of said objective character string, a correct division pattern is selected from said plurality of candidates with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability.

12. (Previously Presented) The character string dividing method in accordance with claim 11, wherein a score of each candidate is calculated when there are a plurality of candidates for a division pattern of said objective character string,

said score is a sum of character joint probabilities at respective division points of said objective character string in accordance with a division pattern of said each candidate, and

a candidate having the smallest score is selected as said correct division pattern.

13. (Previously Presented) The character string dividing method in accordance with claim 11, wherein

a score of each candidate is calculated when there are a plurality of candidates for a division pattern of said objective character string,

said score is a product of character joint probabilities at respective division points of said objective character string in accordance with a division pattern of said each candidate, and

a candidate having the smallest score is selected as said correct division pattern.

14. (Previously Presented) The character string dividing method in accordance with claim 11, wherein

a calculated character joint probability is given to each division point of said candidate;

a constant value is assigned to each point between two characters not divided;

a score of each candidate is calculated based on a sum of said character joint probability and said constant value thus assigned; and

a candidate having the smallest score is selected as said correct division pattern.

15. (Previously Presented) The character string dividing method in accordance with claim 11, wherein a calculated character joint probability is given to each division point of said candidate;

a constant value is assigned to each point between two characters not divided;

a score of each candidate is calculated based on a product of said character joint probability and said constant value thus assigned; and

a candidate having the smallest score is selected as said correct division pattern.

16. (Currently Amended) A character string dividing system for segmenting a character string into a plurality of words, comprising:

input means for receiving a document to be processed for word division;

document data storing means serving as a document database for storing a received document to be division processed;

character joint probability calculating means for calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in said document database, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary;

probability table storing means for storing a table of calculated character joint probabilities;

word dictionary storing means for storing a word dictionary prepared or produced beforehand;

unknown word estimating means for estimating unknown words not registered in said word dictionary;

division pattern producing means for producing a plurality of candidates for a division pattern of an objective character string with reference to information of said word dictionary and said estimated unknown words;

correct pattern selecting means for selecting a correct division pattern from said plurality of candidates with reference to said table of character joint probabilities; and

output means for outputting said selected correct division pattern as a division result of said objective character string.

17. (Currently Amended) A character string dividing method for segmenting a character string into a plurality of words, said method comprising the steps of:

calculating substantially automatically without human tagging from documents to be processed a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database to be processed for word division, in which calculation of said character joint probability is performed based only on the information involved in said given document database without referring to any dictionary;

storing calculated character joint probabilities; and

segmenting an objective character string into a plurality of words with reference to dictionary words and estimated unknown words,

wherein, when there are a plurality of candidates for a division pattern of said objective character string, a correct division pattern is selected from said plurality of candidates with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability.

18. (Original) The character string dividing method in accordance with claim 17, wherein it is checked if any word starts from a certain character position (i) when a preceding word ends at a character position (i-1) and, when no dictionary word starting from said character position (i) is present, appropriate character strings are added as unknown words starting from said character position (i), where said character strings to be added have a character length not smaller than n and not larger than m, where n and m are positive integers.

19. (Original) The character string dividing method in accordance with claim 17, wherein

a constant value given to said unknown word is larger than a constant value given to said dictionary word, a score of each candidate is calculated based on a sum of said constant values given to said unknown word and said dictionary word in addition to a sum of calculated joint probabilities at respective division points, and

a candidate having the smallest score is selected as said correct division pattern.

20. (Cancelled)

21. (Previously Presented) A character string dividing method for segmenting a character string into a plurality of words, said method comprising:

calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database;

storing calculated character joint probabilities; and

segmenting an objective character string into a plurality of words with reference to dictionary words and estimated unknown words, wherein,

(a) when there are a plurality of candidates for a division pattern of said objective character string, a correct division pattern is selected from said plurality of candidates with reference to calculated character joint probabilities so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability,

(b) a constant value given to said unknown word is larger than a constant value given to said dictionary word,

(c) a score of each candidate is calculated based on a product of said constant values given to said unknown word and said dictionary word in addition to a product of calculated joint probabilities at respective division points, and

(d) a candidate having the smallest score is selected as said correct division pattern.